Insect Molecular Biology (2003) **12**(4), 383–392

Fat body expressed yolk protein genes in *Hyphantria cunea* are related to the YP4 follicular epithelium yolk protein subunit gene of pyralid moths

H. M. Cheon*, H. J. Kim*, C. Y. Yun†, H. J. Lee‡, I. H. Lee§, P. D. Shirk¶ and S. J. Seo*

**Division of Life Science, College of Natural Sciences, Gyeongsang National University, Jinju, Korea; †Department of Biology, Daejeon University, Daejeon, Korea; ‡Department of Biology, Korea University, Seoul, Korea; §Department of Life Science, Hoseo University, Asan, Korea; ¶USDA ARS, Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, Florida, USA*

Abstract

cDNA clones for two of the yolk proteins, YP1 and YP2, produced by the fat body of the moth, *Hyphantria cunea*, were sequenced and found to be homologous to the follicular epithelium yolk proteins of pyralid moths. Both cDNA clones coded for polypeptides of 290 residues and the deduced amino acid sequence identity between YP1 and YP2 was very high (79.0%). Analysis of the secondary structure of the predicted polypeptides suggests that YP1 and YP2 do not form heteromeric proteins because of differences in secondary structure due to the lack of alpha helices in YP1. Northern blot analysis showed that the transcripts for YP1 (1.2 kb) and YP2 (1.1 kb) were present primarily in the female fat body with only trace levels detectable in the ovary of the adult female. In a developmental study, the YP1 and YP2 transcripts were first detectable in 10-day-old pupae and increased into the adult stage. These results suggest that the YP1 and YP2 genes in *H. cunea* have been recruited to replace the vitellogenin gene as the primary source of yolk proteins. During this process they have acquired a modified pattern of expression that is different from homologous genes reported in pyralid moths. The assessment of the evolution of proteinaceous yolk in these moths should serve as an excellent model for the evolution of gene recruitment.

Keywords: *Hyphantria cunea*, yolk protein, vitellogenin, cDNA, RT-PCR.

Introduction

The major yolk proteins found in the eggs of many insect orders have evolved from the vitellogenin gene family present in animal species from round worms to vertebrates (reviewed in Byrne *et al*., 1989; Raikhel & Dhadialla, 1992; Chen *et al*., 1997; Hagedorn *et al*., 1998; Sappington & Raikhel, 1998). Typically, vitellogenin is produced in the fat body of the insect as a large provitellogenin, which is proteolytically cleaved into two subunits that are then glycosylated, phosphorylated and lipidated before secretion into the haemolymph. Vitellogenin is transported in the haemolymph from the fat body to the ovary, where it is taken up by a receptor-mediated process into the oocytes and stored in yolk spheres as vitellin.

The major exception to the utilization of vitellogenins as yolk proteins is the cyclorrhaphan Diptera. The yolk proteins of *Drosophila* and the higher flies have evolved from an acid lipase gene family (Bownes, 1992) and there is no evidence of the vitellogenin gene within the genome (Adams *et al*., 2000). The yolk proteins of cyclorrhaphans are produced in the follicular epithelium cells, and in most cases the fat body as well, and are taken up by the oocytes (reviewed in Sappington & Raikhel, 1998).

Most Lepidoptera studied utilize vitellogenin as the major precursor for their yolk proteins (Kunkel & Nordin, 1986; Chen *et al*., 1997; Sappington & Raikhel, 1998). However, bombycid and pyralid moths also produce significant quantities of proteins in the follicular epithelium, which are taken up by the oocytes during vitellogenesis as important components of the proteinacious yolk (Irie & Yamashita, 1983; Shirk *et al*., 1984; Bean *et al*., 1988; Shirk, 1987). Egg

Received 2 January 2003; accepted after revision 9 April 2003. Correspondence: S. J. Seo, Division of Life Science, College of Natural Sciences, Gyeongsang National University, Jinju 660-701, Korea. Tel.: +82 55 751 5951; fax: +82 55 754 0086; e-mail: sookjae@gshp.gsnu.ac.kr

Note: The sequence reported in this paper has been deposited in the GenBank data base, accession no. AF497847, AF497848 (for the nucleotide sequence of the coding region of the *Hyphantria cunea* yolk protein 1 and 2).

specific protein (ESP) in *Bombyx mori* (Sato & Yamashita, 1991) and the YP2 subunit of the follicular epithelium yolk proteins (FEYP) of *Plodia interpunctella* and *Galleria mellonella* (Shirk & Perera, 1998) share significant sequence similarity with vertebrate lipoprotein lipases (Sato & Yamashita, 1991; Shirk & Perera, 1998; Sappington, 2002). The YP4 subunit of the FEYP of *G. mellonella* and *P. interpunctella* does not show any significant sequence similarity to any other known insect proteins (Rajaratnam, 1996a; Perera & Shirk, 1999).

Unlike most Lepidoptera, the yolk of the fall webworm moth, *Hyphantria cunea*, contains three yolk proteins (Lee *et al*., 1995). Yolk proteins 1–3 consist of polypeptides of 34 kDa (YP1), 35.4 kDa (YP2) and 18 kDa (YP3), respectively. Each of these proteins is produced in the fat body and transported in the haemolymph (Lee *et al*., 1995). These three yolk proteins are found in both the haemolymph and ovaries of late pupae and adults, and together they constitute 40% of total egg protein (Han & Kim, 1986). Both YP1 and YP2 appear in yolk spheres after the onset of patency during vitellogenesis and can be detected until egg hatch (Lee *et al*., 1995; Cheon *et al*., 2001), demonstrating their utilization during embryogenesis. On the other hand, while YP3 accumulates in the yolk sphere concomitantly with YP1 and YP2, it persists through the 2nd instar (data not shown). Along with the three YPs, storage protein 1 also accumulates in the yolk spheres but has a slightly different developmental profile than the YPs. It is synthesized by the follicle cells, and detected in early yolk spheres prior to the onset of patency. Peak accumulation of storage protein 1 occurs during vitellogenesis in 10-day-old pupae (Cheon *et al*., 2001).

To determine the relationship of YP1 and YP2 to other yolk proteins, we constructed and screened a cDNA library representing RNAs from pupal and adult stages. Although they were produced in the fat body, the YP1 and YP2 of *H. cunea* were found to be homologues of the YP4 subunit of the FEYP in pyralid moths.

Results

Identification of yolk proteins from H. cunea

Ovaries accumulated several peptides, three of which were identified as YP1, YP2 and YP3, based on their relative mobilities in native and denaturing SDS/PAGE (Han & Kim, 1986; Lee *et al*., 1995). SDS/PAGE (Fig. 1A) resolved extracts of ovaries from 6-day-old pupae to adults into three polypeptide bands with molecular weights of 34 kDa (YP1), 35.5 kDa (YP2) and 18 kDa (YP3). YP1 and YP2 of *H. cunea* were first detected by Coomassie blue staining in extracts from 8-day-old pupal ovaries. They were present in large quantities in the ovaries by the end of the pupal and early adult stages (Fig. 1A). After oviposition, yolk polypeptides were present at relatively constant levels in the eggs through hatching (Fig. 1B).

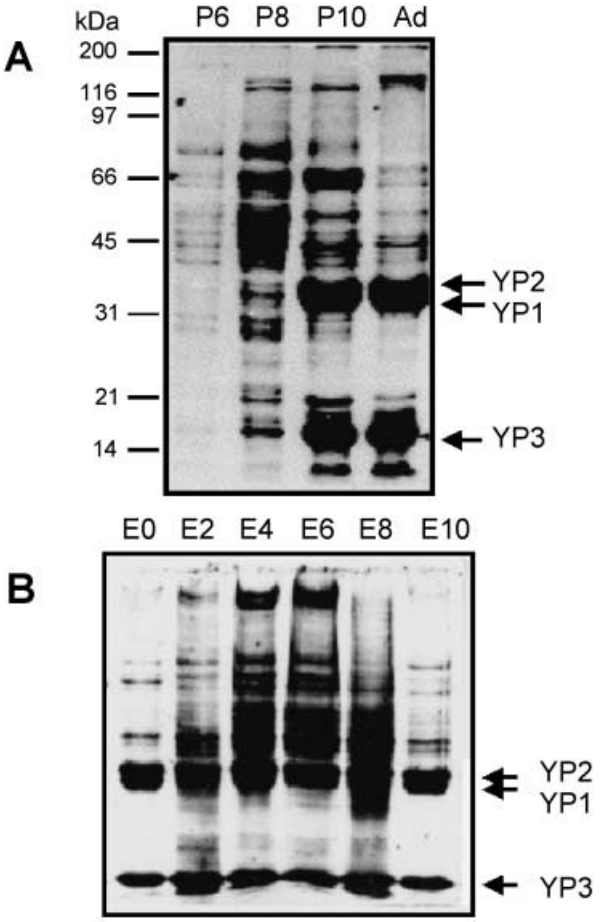

Figure 1. SDS/PAGE of ovary (A) and egg (B) extracts of female *H. cunea*. Each lane was loaded with 30 μg of proteins and the gels were stained with Coomassie blue. Positions of molecular mass standards ($\times 10^3$) are indicated on the left. P6, P8 and P10: 6, 8 and 10-day-old pupae; Ad, newly eclosed adult; E0–E10, eggs 0, 2, 4, 6, 8 and 10 days old after oviposition.

Isolation and characterization of cDNA clones and the predicted amino acid sequences of YP1 and YP2

To isolate cDNA clones of yolk proteins, degenerate PCR primers corresponding to the N-terminal and internal peptide sequences were designed to produce a PCR product (Fig. 2). Thirty-eight N-terminal and nine internal amino acids were sequenced (data not shown), and the sequences were used to design the degenerate primers.

RT-PCR of total RNA isolated from 10-day-old pupal fat body generated ≈ 200 bp PCR products from the primers for YP1 and YP2. The PCR products from RT-PCR were cloned into pGEM-T Easy vector and sequenced. The resulting sequences contained an open reading frame corresponding in part to the respective N-terminal amino acid sequences of YP1 or YP2. These PCR products were used as probes to screen for the homologous cDNA clone for each YP.

The cDNA library constructed from *H. cunea* mRNA was screened with each of the two different ^{32}P-labelled PCR

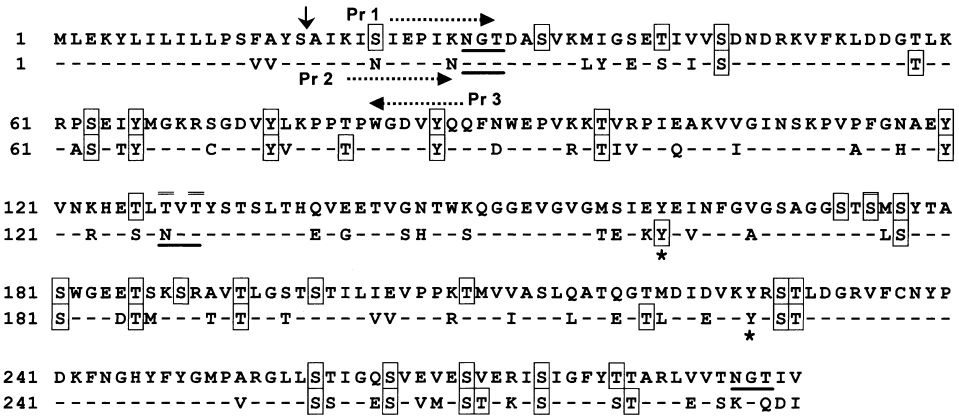

Figure 2. Predicted amino acid sequences from the cDNA clones for *H. cunea* YP1 and YP2. The amino acid differences are indicated, and the site of signal sequence cleavage shown by an arrow. Identical residues are indicated by dashes. The dotted arrows indicate sequences corresponding to the degenerate sense primers (Pr1, YP1; Pr2, YP2) and antisense primer (Pr3 for YP1 and YP2) used for RT-PCR. Underlines and double overlines indicate potential N-glycosylation and O-glycosylation sites, respectively. Potential sulphation sites are marked by asterisks and potential phosphorylation sites are boxed.

Table 1. Amino acid identity matrix of yolk polypeptides and vitellogenins from Lepidopteran insects. Values indicates percentage identical residues in pairwise comparisons of sequences aligned by ClustalV program of DNASTAR (data not shown). Ms mVg, microvitellogenin of *Maduca sexta* (A28068 *Wang et al.*, 1988); GmYP2, yolk protein 2 of *G. mellonella* (AAB0908 *Rajaratnam*, 1996b); Pi YP2, yolk protein 2 of *P. interpunctella* (AAC62229 *Shirk & Perera*, 1998); Bm ESP, egg specific protein of *B. mori* (BAA02091 *Sato & Yamashita*, 1991); Bm Vg, vitellogenin of *B. mori* (S45289 *Yano et al.*, 1994); Ld Vg, vitellogenin of *L. dispar* (T43162 *Hiremath & Lehtoma*, 1997)

Sequence		Percent Identity									
Hc YP1	**1**										
Hc YP2	**2**	79.0									
Ms mVg	**3**	10.8	10.8								
Gm YP2	**4**	11.0	11.7	11.2							
Gm YP4	**5**	31.5	30.4	11.6	11.5						
Pi YP2	**6**	12.4	12.8	13.7	45.6	12.2					
Pi YP4	**7**	33.8	33.1	14.1	12.0	43.4	14.4				
Bm ESP	**8**	11.4	10.7	12.4	38.1	11.9	40.6	11.7			
Bm Vg	**9**	12.1	13.8	12.9	12.6	11.2	12.0	11.1			
Ld Vg	**10**	12.8	13.4	12.4	10.7	12.6	11.0	11.7	11.4	41.6	
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

products for YP1 and YP2 as probes. From a total of 50 000 screened plaques, three positive plaques were isolated. Each of the positive phagemids was digested with *Eco*RI and *Xho*I, and was found to contain an insert of approximately 2 kb and 1.5 kb, respectively. The two species of cDNA clones were sequenced and the predicted amino acid sequences were compared with the N-terminal sequences derived from purified *H. cunea* YPs, confirming the identity of the cDNA clones for YP1 (AF497847) and YP2 (AF497848). The complete cDNA sequences for *H. cunea* YP1 and YP2 are shown, extending 1180 bp and 1089 bp, respectively. Each contains a single open reading frame that encodes polypeptides of 290 amino acids (Fig. 2).

The deduced amino acid sequences for YP1 and YP2 indicated the presence of a 17 amino acid signal peptide that was 5′ to the N-terminal sequences of the mature *H. cunea* YPs. The N-terminal sequence of purified YP2 was identical with amino acids 18–55 of the deduced amino acid sequence from the cDNA. The N-terminal sequences of purified YP1 matched the deduced amino acid sequences from the cDNA with exception of two amino acid residues (data not shown). The difference between the two sequences may be due to an error in N-terminal sequencing of the protein. Two potential glycosylation sites (NXT/S) are present in the *H. cunea* YPs, suggesting that they are glycoproteins (Han & Kim, 1986). High content of serine and threonine which could serve as potential phosphorylation and O-linked glycosylation sites are found in the *H. cunea* YPs (Fig. 2).

Sequence similarity between Hyphantria *YPs and other Lepidopteran yolk proteins*

Sequence identity between YP1 and YP2 of *H. cunea* was significant (79.0%) (Table 1). When conservative substitutions are considered, the overall sequence similarity is 90.0% (data not shown). In addition, the glycine and proline residues, which might be involved in breaks or turns in the

```
  1  MLEKYLILILLPSFAYSAIKISI-EPIKNGTDASVKMI--GSETIVVSDNDRKVFKLDDG  HcYP1
  1  MLEKYLILILLPSVVYSAIKINI-EPINNGTDASVKLY--GEESIIVSDNDRKVFKLDDG  HcYP2
  1  MASLYL LLVLPALVLARIDVQLSGEFNDDSHNNLKVYYSGSQAGVITDAERTTFGLSDA  PiYP4
  1  MVSKYLAILLFIPIIYAKIQVNVVAS ENEAETGEPGW---KNVDIITDNERYTFQLTDN  GmYP4
     *   **          *                          *       * *
 58  TLKRPSEIYMGKRSGDVYLKPPTPWGDVYQQFNWEPVKKTVRPIEAKVVGINSKPVPFGN  HcYP1
 58  TLKRASETYMGKRCGDVYVKPPTPWGDVYQQFDWEPVRKTIVPIQAKVIGINSKPVAFGH  HcYP2
 60  ILKDSIGAYFGRRPDDAYLRSPTPWGDLYSTFGWDQVLRTLVPKNGKILGITSQPMIITK  PiYP4
 57  NLKNAVQSYFGQRPDDAFLRSPTPWGDLYQVYGWPQVARSLSSSESKYLQVSSKPSIILT  GmYP4
      **      * * *   *      ****** *     *   *          *      * *
118  AEYVNKHETLTVTYSTSLTHQVEETVGNTWKQGGEVGVGMSIEYEINFGVGSAGGSTSMS  HcYP1
118  AEYVNRHESLNVTYSTSLTHEVGETVSHTWSQGGEVGVGTEIKYEVNFGAGSAGGSTSLS  HcYP2
120  QLF-ENNSSKPATFDVGISQSVQNTVKSSWSQGGSLNVAQEINYGFNIEVIKGGGKTAFS  PiYP4
117  QHF-RNNSTQPATFKAQIQQQVQNTVTSTWEKGGELTVGQEIEYGFDIKVVSVGGKTSFS  GmYP4
                  *         *   **    *   **    *    * *          ** *  *
178  YTASWGEETSKSRAVTLGSTSTILIEVPPKTMVVASLQATQGTMDIDVKYRSTLDGRVFC  HcYP1
177  YTASWGEDTMKSRTVTLGTTSTILVVVPPRTMVIASLLATEGTLDIEVKYRSTLDGRVFC  HcYP2
179  YTSTWGRNTEKSESVTIGSQSGMKILLQPGQAVVAQLQATKGTIRMEVEYEASLSGASAV  PiYP4
176  YTSRWGESVSKSETVTVGSESGVEITLEPGQAIIAELLATRGTMEIQVDYEATLSGATAV  GmYP4
     **  **    **  ** *  *       *      * * ** **    * *   * *
238  NYPDKFNGHY-FYGMP------ARGLLSTIGQSVEVESVERISIGFYTTARLVVTNGTI-  HcYP1
237  NYPDKFNGHY-FYGMP------VRGLLSSIGESVVMESTEKISIGFYSTARLEVSKGQD-  HcYP2
239  NYDDGYRGHH-FWSL------DIRAIMASVGLPNQKVSKEVIEIDFYSQSRVVVNDKSTN  PiYP4
236  NYANTFKGHHHFWASGHQCCDDVWRLEPHRSLPGS----------Y QARLLLQLSRCD  GmYP4
     **      **  *                                  *   *
290  -------V                                                      HcYP1
289  -------I                                                      HcYP2
292  IKLMEIGF                                                      PiYP4
284  QR-----C                                                      GmYP4
```

Figure 3. Multiple alignments of the amino acid sequences of YP1 and YP2 from *H. cunea* and related amino acid sequences of YP4 of *P. interpunctella* (PiYP4, AAD13187, Perera & Shirk, 1999) and *G. mellonella* (GmYP4, AAB41454, Rajaratnam, 1996a). A dash (–) indicates an alignment gap. Asterisks indicate amino acids that are identical among the four yolk proteins.

secondary structure, are well-conserved (Fig. 3). A BLAST search was made for each of the two *H. cunea* cDNA sequences and their predicted amino acid sequences. Multiple alignments revealed significant sequence similarity in the predicted amino acid sequences of *H. cunea* YP1 and YP2 with the FEYP subunit YP4 from *P. interpunctella* (33.1–33.8% identity) and *G. mellonella* (30.4–31.5% identity). Whereas YP1 and YP2 of *H. cunea* did not share sequence similarity with vitellogenin from three species; ESP from *B. mori*, and YP2 from *G. mellonella* and *P. interpunctella* (10.8–13.8%) (Table 1).

The high degree of similarity between YP1 and YP2 from *H. cunea* raises the question as to how the two native yolk proteins are comprised of only homomeric units (Lee, 1993; Lee *et al*., 1995). In order to gain some perspective on this, the predicted amino acid sequences for the four closely related YPs from *H. cunea*, *G. mellonella*, and *P. interpunctella* were analysed for structural patterns using hydrophilicity plots (Kyte & Doolittle, 1982) (Fig. 4A) and Robson–Garnier 2y structural plots (Fig. 4B). All of the proteins showed a similar hydrophilicity pattern across the entire sequence with most of the protein consisting of large hydrophilic regions (Fig. 4A). *H. cunea* YP1 has no predicted tendency to form helices whereas *H. cunea* YP2 is predicted to contain substantial amount of helices (Fig. 4B). This suggests that even though there is high amino acid similarity between YP1 and YP2, the differences may be sufficient to create folding patterns that preclude formation of heteromeric proteins. *P. interpunctella* YP4 is predicted to have extended helical regions similar to *H. cunea* YP2 while *G. mellonella* YP4 has fewer predicted helices.

Stage- and sex-specificity of YP mRNA accumulation

The profiles of developmental and sex-specific expression for YP1 and YP2 in *H. cunea* were examined by Northern blot analysis using labelled cDNA as probes. Total RNA was isolated from female fat body during larval, pupal and adult development, and from selected tissues of both sexes. Single transcripts of approximately 1.2 kb for YP1 and 1.1 kb for YP2 were observed (Figs 5 and 6).

In the female fat body, the transcripts for YP1 and YP2 were first detectable in 10-day-old pupae and dramatically increased in the adult stage (Fig. 5), which is consistent with the temporal appearance of the YPs in the ovary (Fig. 1A). Transcripts for YP1 and YP2 were not detected in adult male fat body or testes (Fig. 6). After oviposition, no transcript for either YP was present in the eggs (data not shown). Although small amounts of the transcripts were detectable in the ovary (Fig. 6), the fat body contained much greater amounts of transcript per μg of RNA and appears to be the main site of YP synthesis.

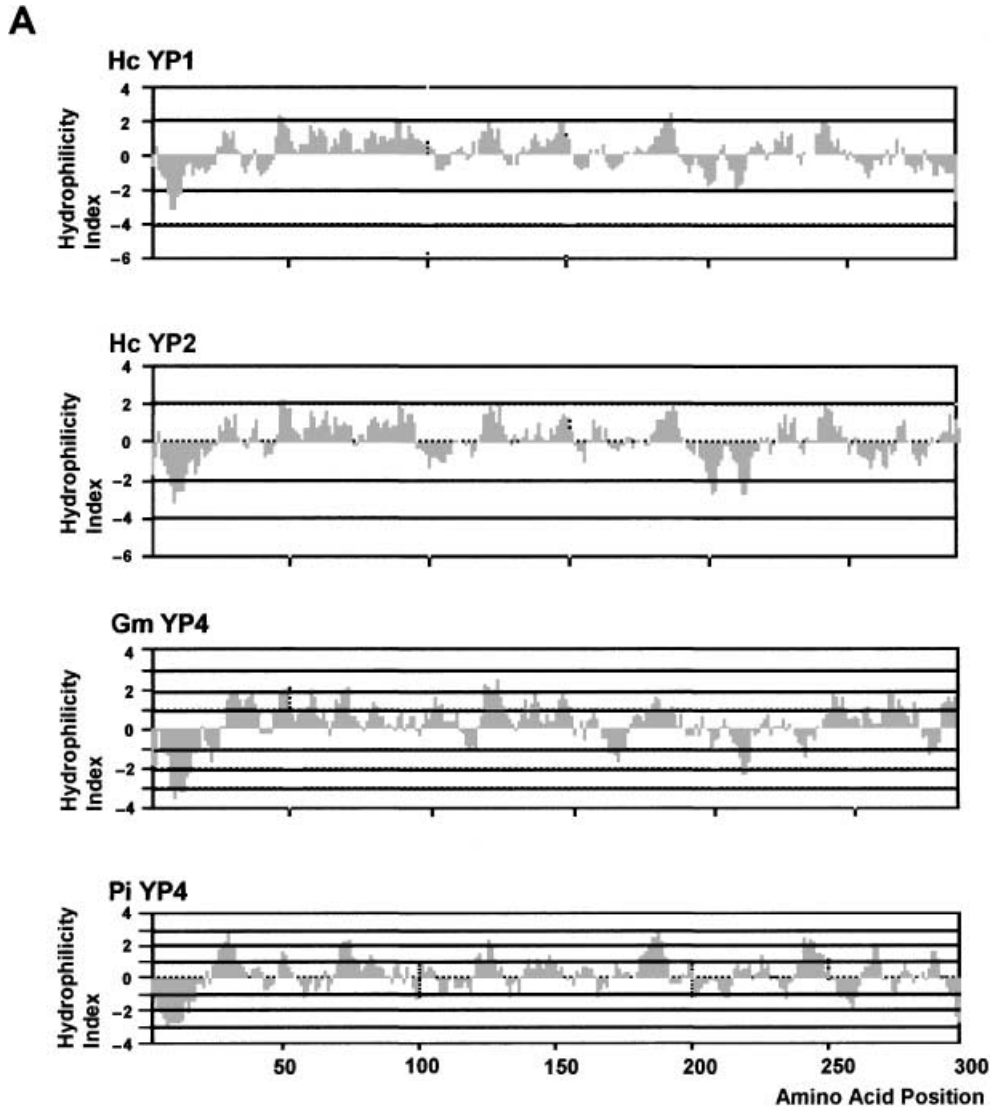

Figure 4. Comparison of hydrophilicity and secondary structure for the predicted amino acid sequences of YP1 (Hc YP1) and YP2 (Hc YP2) from *H. cunea* and YP4 from *G. mellonella* (Gm YP4) and *P. interpunctella* (Pi YP4). (A) Kyte–Doolittle hydrophilicity plots; (B) Robson–Garnier 2y structural plots. Helices are represented in red; sheets in green; and turns in blue.

Discussion

The three major yolk proteins of *H. cunea*, YP1, YP2 and YP3, are small proteins of molecular masses between 18 and 32.2 kDa (Lee *et al.*, 1995). We isolated the YP1 and YP2 cDNA clones containing inserts of 1.2 kb and 1.1 kb, respectively. Each cDNA clone contained a single open reading frame encoding a predicted protein of 290 amino acid residues, and with a calculated molecular mass of 31.8 kDa for the mature polypeptide. The apparent molecular masses of YPs from egg extracts based on SDS/PAGE were approximately 34 kDa (YP1) and 35.5 kDa (YP2), respectively. The difference in their molecular masses could be accounted for by post-translational modifications such as glycosylation and phosphorylation.

Alignment of the predicted amino acid sequences for YP1 and YP2 of *H. cunea* showed they shared a high level of identity with each other. They also shared considerable similarity with the FEYP subunit YP4 from *P. interpunctella* (Perera & Shirk, 1999) and *G. mellonella* (Rajaratnam, 1996a). As with *P. interpunctella* YP4 and *G. mellonella* YP4, both YPs of *H. cunea* share a low level of similarity with spherulin 2a from the slime mould (16.6% and 19%, respectively). Even though there was some similarity between the YP4 sequences from these two moths and spherulin 2a, analysis of the sequences with the Kyte–Doolittle hydrophilicity plots revealed no apparent conservation of

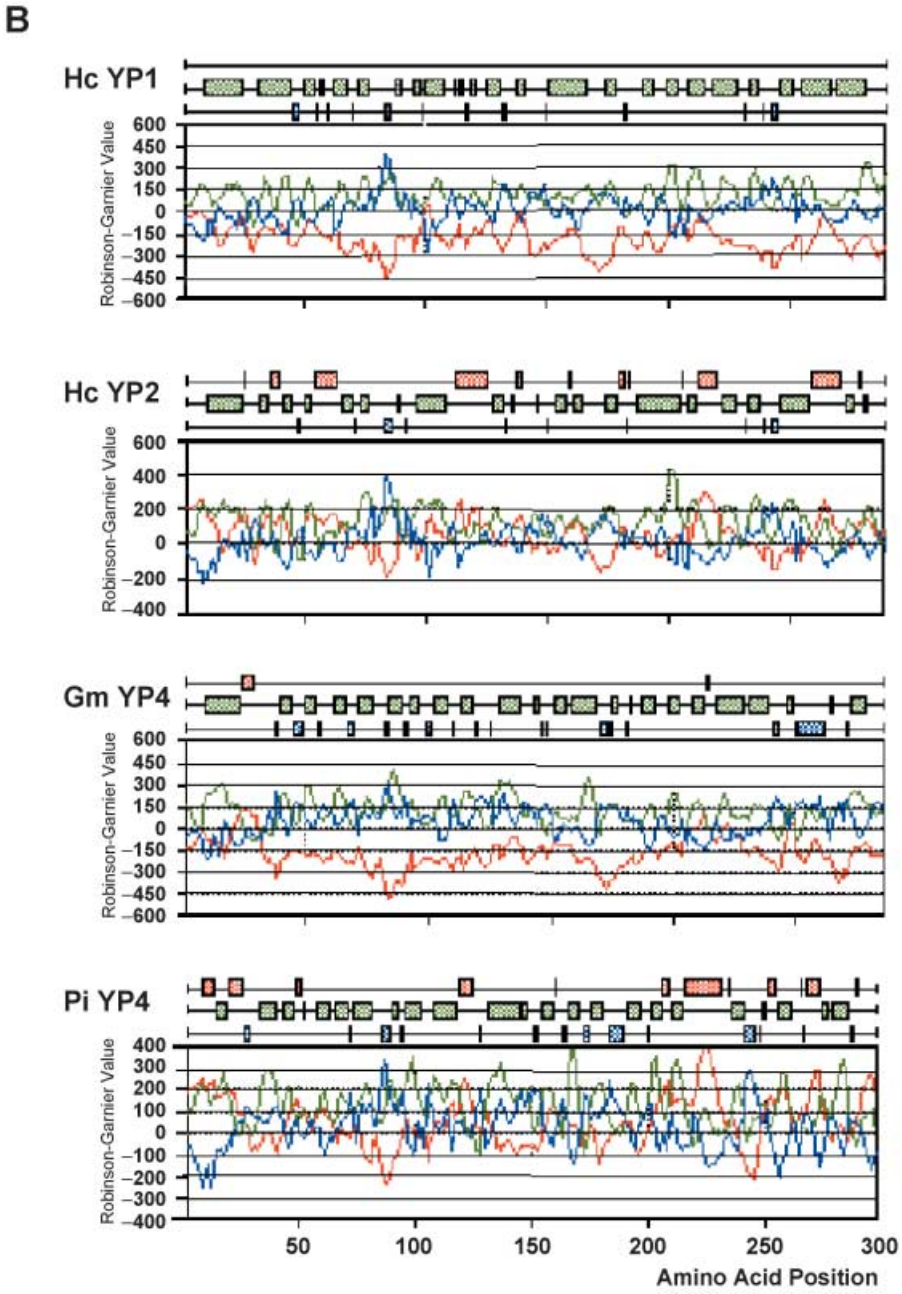

Figure 4. (*continued*)

hydrophilic/hydrophobic regions between any of the three proteins (Fig. 4) (Perera & Shirk, 1999).

YP1 and YP2 of *H. cunea* did not share sequence similarity with any other major yolk proteins such as the lipase-related proteins ESP from *B. mori* (Sato & Yamashita, 1991), the FEYP subunit YP2 from *P. interpunctella* and *G. mellonella* (Shirk & Perera, 1998) and the YPs of *Drosophila melanogaster* and other cyclorrhaphan flies (Sappington, 2002), or with the vitellogenins from other insects (Chen *et al*., 1997; Sappington & Raikhel, 1998).

The lack of an expressed vitellogenin gene in *H. cunea* is similar to the situation in the cyclorrhaphan Diptera. If *Drosophila* is representative of the higher Diptera, it is apparent that the gene for vitellogenin is not present (Adams *et al*., 2000) and that other proteins that accumulate in the oocytes have supplanted the function of vitellin, the form of internalized vitellogenin, as the major yolk protein. For the cyclorrhaphan Diptera and pyralid moths, a class of lipoprotein lipases filled the physiological role of vitellins and evolved into the YPs (Terpstra & Ab, 1988; Bownes,

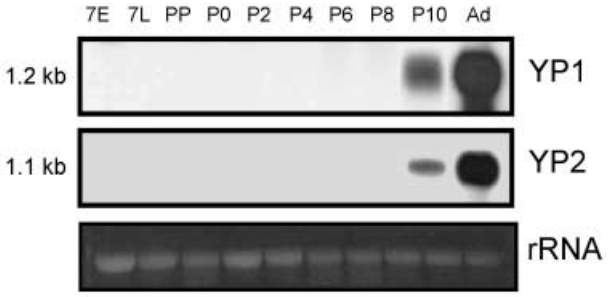

Figure 5. Developmental changes of YP1 and YP2 transcripts in fat body from female *H. cunea*. Total RNA (10 μg) was applied to each lane and subjected to Northern blot analysis. 7E, and 7L, early and late 7th instar larvae; PP, prepupae; P0–P10, pupae at days 0–10; Ad, adult. The lower panel shows ribosomal RNA (rRNA) on the same filter previous to hybridization as a control for the amounts of RNA loaded.

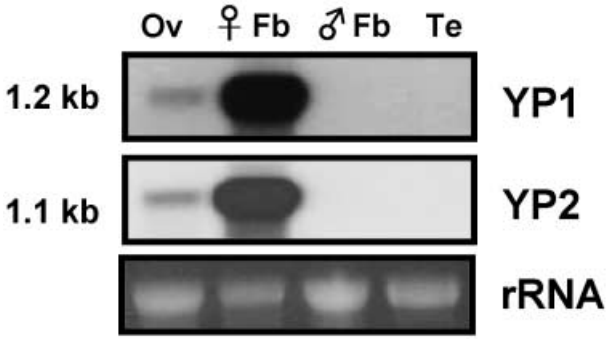

Figure 6. Northern blot analysis of YP1 and YP2 transcripts in mRNA from gonads and fat body of adult female and male *H. cunea*. Total RNA from 10-day-old pupal female ovary (Ov, 30 μg) and fat body (♀Fb, 10 μg), 10 day-old pupal male fat body (♂Fb, 10 μg) and testis (Te, 30 μg) was resolved in denaturing gel conditions and then blotted to a membrane. The blot was hybridized to labelled probe from the cDNA of YP1 and YP2, respectively. The lower panel shows ribosomal RNA (rRNA) on the same filter previous to hybridization as a control for the amounts of RNA loaded.

1992; Sappington, 2002). It is most likely that the YPs were first expressed solely in the follicular epithelium as in the Caribbean fruit fly, *Anastrepha suspensa* (Handler, 1997) and subsequently acquired the promoters that permitted expression in the fat body (Bownes, 1986). The follicle cell expression of lipase-related proteins that are accumulated in the oocytes is also observed in Lepidoptera, even though vitellin is still the major component of the yolk in these moths (Irie & Yamashita, 1983; Shirk *et al*., 1984; Bean *et al*., 1988).

The ESP of *B. mori* (Sato & Yamashita, 1991) and YP2 from *P. interpunctella* and *G. mellonella* (Shirk & Perera, 1998) all share similarity with the vertebrate lipoprotein lipases as well as higher Dipteran YPs (Sappington, 2002), and are expressed in the follicular epithelium during the vitellogenic phase of oocyte maturation. In the case of *H. cunea*, other proteins produced in the follicle cells have been recruited to replace vitellin or have supplanted them as the major yolk proteins. YP1 and YP2 of *H. cunea* are clearly related to the YP4 FEYPs of *G. mellonella* (Rajaratnam, 1996a) and *P. interpunctella* (Perera & Shirk, 1999) which are expressed in the follicular epithelium. However, while small amounts of transcripts were found in the ovaries, YP1 and YP2 were expressed primarily in the fat body of the pharate and adult females, suggesting that they have secondarily acquired promoters necessary for expression in this tissue, as did the YPs of cyclorrhaphan flies. The answer to whether a pseudogene (Nose *et al*., 1997) for vitellogenin remains in the genome of *H. cunea* will require genomic sequence analysis.

The assumption of the physiological role of vitellin by YP1 and YP2 may have been facilitated by the ancestral functional activities of these proteins. The deduced amino acid sequences of YP1 and YP2 contain a region between amino acids 169 and 199 composed of 50% serine and threonine. These residues could serve as potential phosphorylation (Blom *et al*., 1999) and O-linked glycosylation sites. Regions with high serine and threonine content are also present in *B. mori* ESP (Sato & Yamashita, 1991), *B. mori* vitellogenin (Yano *et al*., 1994), *Lymantria dispar* vitellogenin (Hiremath & Lehtoma, 1997), and YP4 from *P. interpunctella* (Perera & Shirk, 1999). Whether these regions are important to the function of these proteins might be established by dephosphorylation experiments. Dephosphorylation of chicken and mosquito vitellogenin reduced their uptake by oocytes (Miller *et al*., 1982; Dhadialla *et al*., 1992), indicating that phosphorylated residues may play a role in the maintenance of vitellogenin activity or tertiary structure that is critical for binding between vitellogenin and its receptor.

YP1 and YP2 from *H. cunea* share 79.0% similarity between their amino acid sequences. This suggests that the two YPs may have arisen by a recent gene duplication. However, even though the two proteins share such a high level of similarity in amino acid sequence, they do not associate to form a heteromeric protein (Lee, 1993; Lee *et al*., 1995). The divergence in their sequences must have led to significant structural changes that do not allow the two proteins to bind with each other. Determining whether the two proteins share the same receptor for uptake into the oocyte will require the isolation and characterization of the receptors.

From these observations, it is apparent that the evolution of yolk proteins in the Lepidoptera offers a unique opportunity to assess the mechanisms that lead to functional recruitment and/or replacement of proteins. The accumulation of vitellin as the major protein in the yolk spheres represents the ancestral type (Byrne *et al*., 1989). The modification of this pattern through the accumulation of large amounts of proteins produced in the follicular epithelium as in bombycid and pyralid moths, or the replacement of vitellin by other proteins, as in *H. cunea*, reflects the

evolutionary flexibility of this process. Analysis of yolk proteins accumulated in oocytes in species with intermediate patterns should provide a much better understanding of how the divergence occurred.

Experimental procedures

Animals

Fall webworms, *Hyphantria cunea*, were obtained from a colony in the Laboratory of Insect Conservation, Department of Sericulture and Entomology, National Institute Agricultural Science and Technology, Suwon, Korea. They were reared on an artificial diet (Ito & Tanaka, 1960) at 27 °C and 75% relative humidity with a photoperiod of 16 L : 8 D. Under the laboratory conditions, eggs hatch in 10 days and the pupal stage lasts 10 days.

Electrophoresis of ovary and egg proteins

Ovaries were dissected from female pupae and adults and cleaned of attached fat body cells thoroughly in Ringer's solution (150 mM NaCl, 1.8 mM $CaCl_2$, 1.3 mM KCl, 10 mM Tris, pH 7.4). Developing ovaries and eggs were homogenized in five volumes of Ringer's solutions using a Dounce glass/Teflon homogenizer. The homogenates were centrifuged at 10 000 ***g*** for 30 min at 4 °C. The supernatant was removed and stored at −70 °C. SDS/PAGE of the soluble proteins in the supernatant of samples was performed according to the method of Laemmli (1970) using 12.5% (w/v) gels. Samples were prepared for electrophoresis by heating at 90 °C for 9 min in the presence of 2% (w/v) SDS and 5% (v/v) 2-mercaptoethanol. Gels were stained with Coomassie blue after electrophoresis.

Protein sequencing

YP1 and YP2 were purified from the adult ovaries using KBr ultracentrifugation and DEAE according to Lee *et al.* (1995). The amino acid sequences for the N-termini of purified YPs were determined using an Applied Biosystems 476A Sequencer (Biotechnology Laboratory, University of British Columbia, Vancouver, Canada). For internal peptide sequencing, YPs were digested with trypsin, chymotrypsin and endoproteinase Lys-C. The amino acid sequences for internal peptide fragments were determined using an Applied Biosystems Procise 491 (Korea Basic Science Institute, Korea).

Isolation of RNA

Total RNA was isolated from tissues using the RNAEasy mini kit (Qiagen Inc. Chatsworth, US) according to the protocol recommended by the manufacturer. All RNA samples were evaluated in agarose gels to ensure that they contained intact rRNA and were free of contaminating DNA.

PCR

Degenerate primers were designed on the basis of the amino acid sequences derived from mature and fragmented peptides of purified YP1 and YP2. The primers for each peptide are listed below with the numbers in parentheses referring to amino acid positions as given in Fig. 2.

YP1 forward primer (23–29 amino acids): 5′-ATHGARCCNATH-AARAARGGN-3′

YP1 reverse primer (82–87 amino acids): 5′-YTGRTANACRTC-NCCCCA-3′

YP2 forward primer (20–27 amino acids): 5′-AARATHAAYATH-GARCCNATHAAY-3′

YP2 reverse primer (82–87 amino acids): 5′-YTGRTANACRTC-NCCCCA-3′

PCR amplification was conducted in a programmable thermal controller (MJ Research, Watertown, MA) using template cDNA reverse-transcribed from 5 g of total RNA prepared from the ovaries of 10-day-old pupae. The PCR cycles were as follows: 5 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 52 °C, 1 min at 72 °C; 7 min at 72 °C.

The PCR products for each set of primers were separated on a 1% (w/v) agarose gel and a 0.2 kb fragment was obtained for each reaction. These fragments were excised from the gel, eluted, ligated into a T-vector, and amplified in XL1 Blue competent cells.

cDNA library screening

A cDNA library was constructed in Lambda-ZAP using the ZAP cDNA synthesis kit (Stratagene, CA, US) and 5 μg of poly(A)+ RNA isolated from the whole bodies of all pupal stages and adults. For screening, 50 000 plaques were plated on 15-cm diameter agar plates. Nitrocellulose filter (Schleicher & Schuell, Dassel, Germany) blots were taken from the plates and hybridized at high stringency (65 °C and 0.1× SSC) with the ^{32}P-labelled PCR product. Positive clones were rescreened at low density on 9-cm diameter agar plates. Colony hybridization yielded several positive clones from the library. Positive plaques were individually isolated in phage buffer by the plate lysate method followed by lambda DNA extraction (Sambrook *et al.*, 1989). Inserts of the positive clones were subcloned into the unique *Eco*RI site of pBluescript KS(+).

Subcloning and DNA sequencing

cDNA fragments that hybridized with the radiolabelled PCR product were removed from the gels by electroelution, followed by phenol extraction and ethanol precipitation in the presence of 2 μg glycogen. Subsequently, these DNA fragments were ligated into pBluescript KS(+), followed by transformation into JM109 cells (Sambrook *et al.*, 1989).

The sequencing reaction was based on the dideoxynucleotide chain termination method of Sanger *et al.* (1977) using [^{35}S]dATP and the Sequenase version 2.0 polymerase kit (USB Inc.). Template-specific and universal primers derived from pBluescript were used in the sequence reactions. Subclones for sequencing were prepared by ligating suitable restriction fragments into pBluescript followed by transformation into JM109 cells.

Computer-assisted analysis of sequence data

The EMBL DataBank was searched with FASTA and BLAST. Editing and analysis of the DNA sequence data were performed with DNASTAR software (DNASTAR Inc., Wisconsin). Multiple alignment was conducted using PILEUP. Physical and structural analysis of the predicted amino acid sequences was conducted with MacVector 7.0 (Accelrys, New Jersey).

Northern blot analysis

To determine tissue- and sex-specificity of YP expression, tissues were dissected from fat body and gonads of both sexes. Total RNA

from fat body (10 μg) and other tissues (30 μg) were denatured and subjected to electrophoresis in a 1.2% (w/v) agarose gel containing 2.2 M formaldehyde. Following electrophoresis, gels were rinsed in 10× SSC and transferred to nitrocellulose (Schleicher & Schuell, Dassel, Germany) in 10× SSC. Blots were prehybridized with 1.5× SSPE, 7% (v/v) SDS, 10% (v/v) PEG, 0.1 mg/ml sonicated denatured salmon sperm DNA and 0.25 mg/ml BSA for 4 h at 65 °C. Hybridization was performed for 18 h at 65 °C in the prehybridization buffer with 5×10^5 cpm/ml of ^{32}P-labelled probes, prepared by the method of random priming (Feinberg & Vogelstein, 1983). The filter was washed twice with 1× SSC, 0.1% (v/v) SDS at 65 °C for 15 min, and twice more for 15 min with 0.1× SSC, 0.1% (v/v) SDS at 65 °C before exposure to X-ray film at −70 °C.

Acknowledgements

This research was supported by a grant (R14-2002-056-01001-0) from Korea Science Engineering Foundation and Brain Korea 21 project in Gyeongsang National University.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A.YeJ., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C. (2000) The genome sequence of Drosophila melanogaster. *Science* **287**: 2185–2195.

Bean, D.W., Shirk, P.D. and Brookes, V.J. (1988) Characterization of yolk proteins from the eggs of the Indianmeal moth, *Plodia interpunctella*. *Insect Biochem* **18**: 199–210.

Blom, N., Gammeltoft, S. and Brunak, S. (1999) Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**: 1351–1362.

Bownes, M. (1986) Expression of the genes coding for vitellogenin (yolk protein). *Annu Rev Entomol* **31**: 507–531.

Bownes, M. (1992) Why is there sequence similarity between insect yolk proteins and vertebrate lipases? *J Lipid Res* **33**: 777–790.

Byrne, B.M., Gruber, M. and Ab, G. (1989) The evolution of egg yolk proteins. *Prog Biophys Mol Biol* **53**: 33–69.

Chen, J.S., Sappington, T.W. and Raikhel, A.S. (1997) Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. *J Mol Evol* **44**: 440–451.

Cheon, H.M., Kim, H.J., Chung, D.H., Kim, M.O., Park, J.S., Yun, C.Y. and Seo, S.J. (2001) Local expression and distribution of a storage protein in the ovary of *Hyphantria cunea*. *Arch Insect Biochem Physiol* **48**: 111–120.

Dhadialla, T.S., Hays, A.R. and Raikhel, A.S. (1992) Characterization of the solubilized mosquito vitellogenin receptor. *Insect Biochem Mol Biol* **22**: 803–816.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132**: 6–13.

Hagedorn, H.H., Maddison, D.R. and Tu, Z. (1998) The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. *Adv Insect Physiol* **27**: 335–384.

Han, Y.S. and Kim, H.R. (1986) Yolk proteins of fall webworm, *Hyphantria cunea* Drury. *Korean J Entomol* **16**: 41–48.

Handler, A.M. (1997) Developmental regulation of yolk protein gene expression in *Anastrepha suspense*. *Arch Insect Biochem Physiol* **36**: 25–35.

Hiremath, S. and Lehtoma, K. (1997) Complete nucleotide sequence of the vitellogenin mRNA from the Gypsy moth: Novel arrangement of the subunit encoding regions. *Insect Biochem Molec Biol* **27**: 27–35.

Irie, K. and Yamashita, O. (1983) Egg-specific protein in the silkworm, *Bombyx mori*: purification, properties, localization and titre changes during oogenesis and embryogenesis. *Insect Biochem* **13**: 71–80.

Ito, T. and Tanaka, M. (1960) Rearing of the silkworm on an artificial diet and the segregation of pentamolters. *J Seric Sci Japan* **29**: 191–196.

Kunkel, J.G. and Nordin, J.H. (1986) Yolk proteins. In *Comprehensive Insect Physiology Biochemistry Pharmacology* (G.A. Kerkut and L.I. Gilbert, eds), Vol. 1, pp. 83–111. Pergamon Press, London.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**: 105–132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Lee, S.D. (1993) Purification, synthesis, and properties of yolk proteins in *Hyphantria cunea and Periplaneta americana*. Ph D Thesis, University of Korea, Korea.

Lee, S.D., Lee, S.S. and Kim, H.R. (1995) Purification and characterization of yolk protein-2 from the fall webworm, *Hyphantria cunea* Drury. *Arch Insect Biochem Physiol* **28**: 113–129.

Miller, M.S., Benore-Parsones, M. and White, H.B. (1982) Dephosphorylation of chicken riboflavin-binding protein and phosvitin decreases their uptake by oocytes. *J Biol Chem* **257**: 6818–6824.

Nose, Y., Lee, J.M., Ueno, T., Hatakeyama, M. and Oishi, K. (1997) Cloning of cDNA for vitellogenin of the parasitoid wasp, *Pimpla nipponica* (Hymenoptera: Apocrita: Ichneumonidae): vitellogenin primary structure and evolutionary considerations. *Insect Biochem Mol Biol* **27**: 1047–1056.

Perera, O.P. and Shirk, P.D. (1999) cDNA of YP4, a follicular epithelium yolk protein subunit, in the moth, *Plodia interpunctella*. *Arch Insect Biochem Physiol* **40**: 157–164.

Raikhel, A.S. and Dhadialla, T.S. (1992) Accumulation of yolk proteins in insect oocytes. *Annu Rev Entomol* **37**: 217–251.

Rajaratnam, V.S. (1996a) Isolation, characterization and complete nucleotide sequence of a *Galleria mellonella* cDNA coding for the follicle cell-specific yolk protein, YP4. *Insect Biochem Mol Biol* **26**: 545–555.

Rajaratnam, V.S. (1996b) *Galleria mellonella* YP2 partial cDNA; Direct Submission, GenBank Aceession no. U69881.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467.

Sappington, T.W. (2002) The major yolk proteins of higher Diptera are homologs of a class of minor yolk proteins in Lepidoptera. *J Molec Evol* **55**: 470–475.

Sappington, T.W. and Raikhel, A.S. (1998) Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem Molec Biol* **28**: 277–300.

Sato, Y. and Yamashita, O. (1991) Structure and expression of a gene coding for egg-specific protein in the silk worm, *Bombyx mori*. *Insect Biochem* **21**: 495–505.

Shirk, P.D. (1987) Comparison of yolk production in seven pyralid moth species. *Int J Invert Reprod Dev* **11**: 173–188.

Shirk, P.D., Bean, D., Millemann, A.M. and Brookes, V.J. (1984) Identification, synthesis and characterization of the yolk polypeptides of *Plodia interpunctella*. *J Exp Zool* **232**: 87–98.

Shirk, P.D. and Perera, O.P. (1998) The 5′ coding region of the follicular epithelium yolk polypeptide 2 cDNA in the moth, *Plodia interpunctella*, contains an extended coding region. *Arch Insect Biochem Physiol* **39**: 98–108.

Terpstra, P. and Ab, G. (1988) Homology of *Drosophila* yolk proteins and the triacylglycerol lipase family. *J Mol Biol* **202**: 663–665.

Wang, X.-Y., Cole, K.D. and Law, J.H. (1988) cDNA cloning and deduced amino acid sequence of microvitellogenin, a female specific hemolymph and egg protein from the tobacco hornworm, *Manduca sexta*. *J Biol Chem* **263**: 8851–8855.

Yano, K., Sakurai, M.T., Watabe, S., Izumi, S. and Tomino, S. (1994) Structure and expression of mRNA for vitellogenin in *Bombyx mori*. *Biochi Biophys Acta* **1218**: 1–10.